Probability Foundations

Formal Approach on the Model Side

We conduct an experiment that has a set Ω of possible outcomes.
E. g.:

- Rolling a dice ($\Omega = \{1, 2, 3, 4, 5, 6\}$)
- Arrivals of phone calls ($\Omega = \mathbb{N}_0$)
- Bread roll weights ($\Omega = \mathbb{R}_+$)

Such an outcome is called an **elementary event**.

All possible elementary events are called the **frame of discernment** Ω (or sometimes **universe of discourse**).

The set representation stresses the following facts:

- All possible outcomes are covered by the elements of Ω. (**collectively exhaustive**).
- Every possible outcome is represented by exactly one element of Ω. (**mutual disjoint**).

Events

Often, we are interested in *higher-level* events
(e. g. casting an odd number, arrival of at least 5 phone calls or purchasing a bread roll heavier than 80 grams)

Any subset $A \subseteq \Omega$ is called an **event** which **occurs**, if the outcome $\omega_0 \in \Omega$ of the random experiment lies in A.

Since events are sets, we can define for two events A and B:

- $A \cup B$ occurs if A or B occurs; $A \cap B$ occurs if A and B occurs.
- $\overline{A}$ occurs if A does not occur (i. e., if $\Omega \backslash A$ occurs).
- A and B are *mutually exclusive*, iff $A \cap B = \emptyset$.

Event Algebra

A family of sets $\mathcal{E} = \{E_1, \ldots, E_n\}$ is called an **event algebra** on Ω, if the following conditions hold:

- The **certain event** Ω lies in $\mathcal{E}$.
- If $E \in \mathcal{E}$, then $\overline{E} = \Omega \backslash E \in \mathcal{E}$.
- If E_1 and E_2 lie in $\mathcal{E}$, then $E_1 \cup E_2 \in \mathcal{E}$ and $E_1 \cap E_2 \in \mathcal{E}$.

If Ω is not finite, we require the additional property:

For a series of events $E_i \in \mathcal{E}, i \in \mathbb{N},$ the events $\bigcup_{i=1}^{\infty} E_i$ and $\bigcap_{i=1}^{\infty} E_i$ are also in $\mathcal{E}$.

$\mathcal{E}$ is then called a **σ-algebra**.

Side remarks:

Smallest event algebra: $\mathcal{E} = \{\emptyset, \Omega\}$

Largest event algebra (for finite or countable Ω): $\mathcal{E} = 2^{\Omega} = \{A \subseteq \Omega \mid \mathsf{true}\}$

Probability Function

Given an event algebra $\mathcal{E}$, we would like to assign every event $E \in \mathcal{E}$ its probability with a **probability function** $P : \mathcal{E} \to [0, 1]$.

We require P to satisfy the so-called **Kolmogorov Axioms**:

- $\forall E \in \mathcal{E} : \; 0 \;\leq\; P(E) \;\leq\; 1$
- $P(\Omega) \;=\; 1$
- For pairwise disjoint events $E_1, E_2, \ldots \in \mathcal{E}$ holds:

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) \;=\; \sum_{i=1}^{\infty} P(E_i)$$

From these axioms one can conclude the following list of properties:

- $\forall E \in \mathcal{E} : \; P(\overline{E}) \;=\; 1 - P(E)$
- $P(\emptyset) \;=\; 0$
- If $E_1, E_2 \in \mathcal{E}$ are mutually exclusive, then $P(E_1 \cup E_2) = P(E_1) + P(E_2)$.

Elementary Probabilities and Densities

Question 1: How to calculate P?

Question 2: Are there "default" event algebras?

Idea for question 1: We have to find a way of distributing (thus the notion *distribution*) the unit mass of probability over all elements $\omega \in \Omega$.

- If Ω is finite or countable a **probability mass function** p is used:

$$p: \ \Omega \to [0,1] \quad \text{and} \quad \sum_{\omega \in \Omega} p(\omega) = 1$$

- If Ω is uncountable (i. e., continuous) a **probability density function** f is used:

$$f: \ \Omega \to \mathbb{R} \quad \text{and} \quad \int_{\Omega} f(\omega)\, \mathrm{d}\omega = 1$$

“Default” Event Algebras

Idea for question 2 (“default” event algebras) we have to distinguish again between the cardinalities of Ω:

- Ω finite or countable: $\mathcal{E} = 2^{\Omega}$
- Ω uncountable, e. g. $\Omega = \mathbb{R}$: $\mathcal{E} = \mathcal{B}(\mathbb{R})$

$\mathcal{B}(\mathbb{R})$ is the **Borel Algebra**, i. e., the smallest σ-algebra that contains all closed intervals $[a, b] \subset \mathbb{R}$ with $a < b$.

$\mathcal{B}(\mathbb{R})$ also contains all open intervals and single-item sets.

It is sufficient to note here, that all intervals are contained

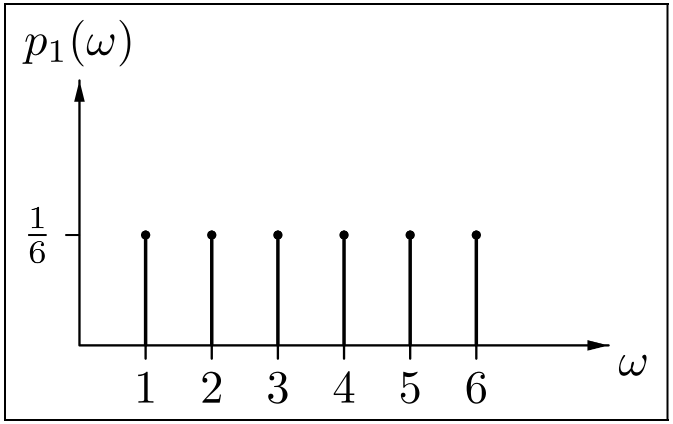
$$\{[a, b]\,,]a, b]\,,]a, b[\,, [a, b[\subset \mathbb{R} \mid a < b\} \subset \mathcal{B}(\mathbb{R})$$

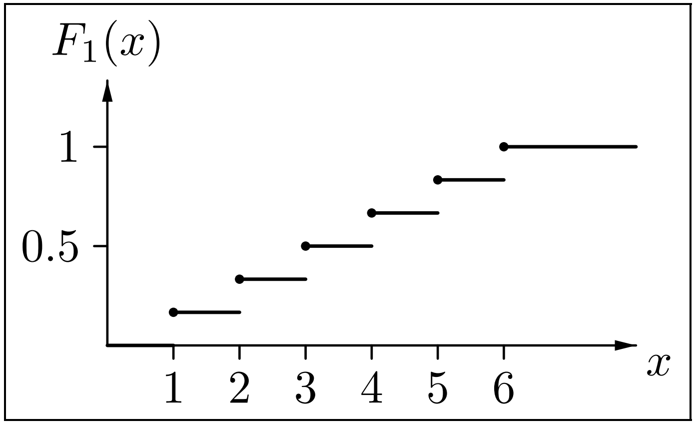
Example: Rolling a Dice

$\Omega = \{1, 2, 3, 4, 5, 6\}$

$$p_1(\omega) = \frac{1}{6}$$

$$F_1(x) = P(X \leq x)$$





$$\sum_{\omega \in \Omega} p_1(\omega) = \sum_{i=1}^{6} p_1(\omega_i) = \sum_{i=1}^{6} \frac{1}{6} = 1$$

$$P(X \leq x) = \sum_{x' \leq x} P(X = x')$$

$$P(a < X \leq b) = F_1(b) - F_1(a)$$

For a random variable X

$$P(X = x) = P(\{X = x\}) = P(X^{-1}(x)) = P(\{\omega \in \Omega \mid X(\omega) = x\})$$

Basics of Applied Probability Theory

Why (Kolmogorov) Axioms?

If P models an *objectively* observable probability, these axioms are obviously reasonable.

However, why should an agent obey formal axioms when modeling degrees of (subjective) belief?

Objective vs. subjective probabilities

Axioms constrain the set of beliefs an agent can abide.

Ramsey (1926) gave a plausible argument why subjective beliefs should respect Kolmogorov axioms:

> The so called "Dutch Book Arguments"

Unconditional Probabilities

$P(A)$ designates the *unconditioned* or *a priori* probability that $A \subseteq \Omega$ occurs if *no* other additional information is present.
For example:

$$P(\mathsf{cavity}) \;=\; 0.1$$

A formally different way to state the same would be via a binary random variable Cavity:

$$P(\mathsf{Cavity} = \mathsf{true}) \;=\; 0.1$$

A priori probabilities are derived from statistical surveys or general rules.

Unconditional Probabilities

In general a random variable can assume more than two values:

$$
\begin{aligned}
P(\ \mathsf{Weather} = \mathsf{sunny}\ ) &= 0.7 \\
P(\ \mathsf{Weather} = \mathsf{rainy}\ ) &= 0.2 \\
P(\ \mathsf{Weather} = \mathsf{cloudy}) &= 0.02 \\
P(\ \mathsf{Weather} = \mathsf{snowy}) &= 0.08 \\
P(\mathsf{Headache} = \mathsf{true}\ ) &= 0.1
\end{aligned}
$$

$P(X)$ designates the vector of probabilities for the (ordered) domain of the random variable X:

$$
\begin{aligned}
P(\mathsf{Weather}) &= \langle 0.7, 0.2, 0.02, 0.08 \rangle \\
P(\mathsf{Headache}) &= \langle 0.1, 0.9 \rangle
\end{aligned}
$$

Both vectors define the respective probability distributions of the two random variables.

Conditional Probabilities

New evidence can alter the probability of an event.

Example: The probability for cavity increases if information about a toothache arises.

With additional information present, the a priori knowledge must not be used!

$P(A \mid B)$ designates the *conditional* or *a posteriori* probability of A *given* the sole observation (*evidence*) B.

$$P(\mathsf{cavity} \mid \mathsf{toothache}) = 0.8$$

For random variables X and Y $P(X \mid Y)$ represents the set of conditional distributions for each possible value of Y.

$P(\mathsf{Weather} \mid \mathsf{Headache})$ consists of the following table:

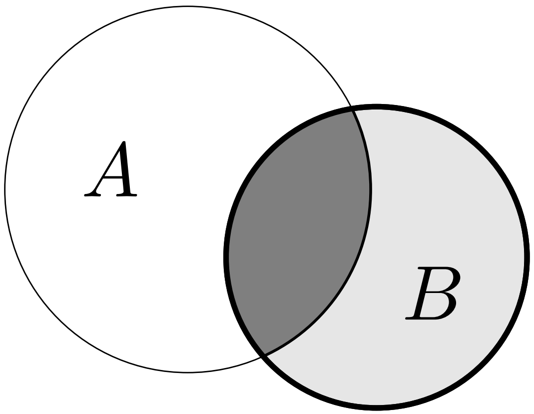
	$\mathsf{h} \mathrel{\widehat{=}} \mathsf{Headache} = \mathsf{true}$	$\neg\mathsf{h} \mathrel{\widehat{=}} \mathsf{Headache} = \mathsf{false}$
$\mathsf{Weather} = \mathsf{sunny}$	$P(\mathsf{W} = \mathsf{sunny} \mid \mathsf{h})$	$P(\mathsf{W} = \mathsf{sunny} \mid \neg\mathsf{h})$
$\mathsf{Weather} = \mathsf{rainy}$	$P(\mathsf{W} = \mathsf{rainy} \mid \mathsf{h})$	$P(\mathsf{W} = \mathsf{rainy} \mid \neg\mathsf{h})$
$\mathsf{Weather} = \mathsf{cloudy}$	$P(\mathsf{W} = \mathsf{cloudy} \mid \mathsf{h})$	$P(\mathsf{W} = \mathsf{cloudy} \mid \neg\mathsf{h})$
$\mathsf{Weather} = \mathsf{snowy}$	$P(\mathsf{W} = \mathsf{snowy} \mid \mathsf{h})$	$P(\mathsf{W} = \mathsf{snowy} \mid \neg\mathsf{h})$

Note that we are dealing with *two* distributions now!
Therefore each column sums up to unity!

Conditional Probabilities

A and B are subsets of Ω

$$\boxed{P(A \mid B) \;=\; \frac{P(A \wedge B)}{P(B)}}$$



Product Rule: $P(A \wedge B) \;=\; P(A \mid B) \cdot P(B)$

Also: $P(A \wedge B) \;=\; P(B \mid A) \cdot P(A)$

A and B are *independent* iff

$$P(A \mid B) \;=\; P(A) \qquad \text{and} \qquad P(B \mid A) \;=\; P(B)$$

Equivalently, iff the following equation holds true:

$$P(A \wedge B) \;=\; P(A) \cdot P(B)$$

Interpretation of Conditional Probabilities

Caution! Common misinterpretation:

"$P(A \mid B) = 0.8$ means, that $P(A) = 0.8$, given B holds."

This statement is wrong.

Joint Probabilities

Let $X_1, \ldots, X_n$ be random variables over the same frame of discernment Ω and event algebra $\mathcal{E}$. Then $\vec{X} = (X_1, \ldots, X_n)$ is called a *random vector* with

$$\vec{X}(\omega) \;=\; (X_1(\omega), \ldots, X_n(\omega))$$

Shorthand notation:

$$P(\vec{X} = (x_1, \ldots, x_n)) \;=\; P(X_1 = x_1, \ldots, X_n = x_n) \;=\; P(x_1, \ldots, x_n)$$

Definition:

$$\begin{aligned} P(X_1 = x_1, \ldots, X_n = x_n) \;&=\; P\left(\left\{\, \omega \in \Omega \;\middle|\; \bigwedge_{i=1}^{n} X_i(\omega) = x_i \,\right\}\right) \\ &=\; P\left(\bigcap_{i=1}^{n} \{X_i = x_i\} \right) \end{aligned}$$

Joint Probabilities

Example: $P(\textsf{Headache}, \textsf{Weather})$ is the *joint probability distribution* of both random variables and consists of the following table:

	$\textsf{h} \mathrel{\widehat{=}} \textsf{Headache} = \textsf{true}$	$\neg\textsf{h} \mathrel{\widehat{=}} \textsf{Headache} = \textsf{false}$
$\textsf{Weather} = \textsf{sunny}$	$P(\textsf{W} = \textsf{sunny} \wedge \textsf{h})$	$P(\textsf{W} = \textsf{sunny} \wedge \neg\textsf{h})$
$\textsf{Weather} = \textsf{rainy}$	$P(\textsf{W} = \textsf{rainy} \wedge \textsf{h})$	$P(\textsf{W} = \textsf{rainy} \wedge \neg\textsf{h})$
$\textsf{Weather} = \textsf{cloudy}$	$P(\textsf{W} = \textsf{cloudy} \wedge \textsf{h})$	$P(\textsf{W} = \textsf{cloudy} \wedge \neg\textsf{h})$
$\textsf{Weather} = \textsf{snowy}$	$P(\textsf{W} = \textsf{snowy} \wedge \textsf{h})$	$P(\textsf{W} = \textsf{snowy} \wedge \neg\textsf{h})$

All table cells sum up to unity.

Calculating with Joint Probabilities

In the finite case all desired probabilities can be computed from a joint probability distribution.

	toothache	¬toothache
cavity	0.04	0.06
¬cavity	0.01	0.89

Example:
$$\begin{aligned} P(\text{cavity} \lor \text{toothache}) &= P(\text{cavity} \land \text{toothache}) \\ &+ P(\neg\text{cavity} \land \text{toothache}) \\ &+ P(\text{cavity} \land \neg\text{toothache}) = 0.11 \end{aligned}$$

Marginalizations:
$$\begin{aligned} \mathrm{P}(\text{cavity}) &= P(\text{cavity} \land \text{toothache}) \\ &+ P(\text{cavity} \land \neg\text{toothache}) = 0.10 \end{aligned}$$

Conditioning:
$$P(\text{cavity} \mid \text{toothache}) = \frac{P(\text{cavity} \land \text{toothache})}{P(\text{toothache})} = \frac{0.04}{0.04 + 0.01} = 0.80$$

Problems

Easiness of computing all desired probabilities comes at an unaffordable price:

Given n random variables with k possible values each, the joint probability distribution contains k^n entries which is infeasible in practical applications.

Hard to handle.

Hard to estimate.

Therefore:

1. Is there a more *dense* representation of joint probability distributions?

2. Is there a more *efficient* way of processing this representation?

The answer is *no* for the general case, however, certain dependencies and independencies can be exploited to reduce the number of parameters to a practical size.

Stochastic Independence

Two events A and B are called *stochastically independent* iff

$$P(A \wedge B) \ = \ P(A) \cdot P(B)$$
$$\Leftrightarrow$$
$$P(A \mid B) \ = \ P(A) \ = \ P(A \mid \overline{B})$$

Two random variables X and Y are *stochastically independent* iff

$$\begin{array}{lrcl} \forall x \in \mathrm{dom}(X) : \forall y \in \mathrm{dom}(Y) : & P(X = x, Y = y) & = & P(X = x) \cdot P(Y = y) \\ & \Leftrightarrow & & \\ \forall x \in \mathrm{dom}(X) : \forall y \in \mathrm{dom}(Y) : & P(X = x \mid Y = y) & = & P(X = x) \end{array}$$

Shorthand notation: $P(X, Y) = P(X) \cdot P(Y)$.
Note the formal difference between $P(A) \in [0, 1]$ and $P(X) \in [0, 1]^{|\mathrm{dom}(X)|}$.

Conditional Independence

Let X, Y and Z be three random variables. We call *X and Y conditionally independent given Z*, iff the following condition holds:

$$\forall x \in \mathrm{dom}(X) : \forall y \in \mathrm{dom}(Y) : \forall z \in \mathrm{dom}(Z) :$$
$$P(X = x, Y = y \mid Z = z) \ = \ P(X = x \mid Z = z) \cdot P(Y = y \mid Z = z)$$

Shorthand notation: $X \perp\!\!\!\perp_P Y \mid Z$

Let $\boldsymbol{X} = \{A_1, \ldots, A_k\}$, $\boldsymbol{Y} = \{B_1, \ldots, B_l\}$ and $\boldsymbol{Z} = \{C_1, \ldots, C_m\}$ be three disjoint sets of random variables. We call *$\boldsymbol{X}$ and $\boldsymbol{Y}$ conditionally independent given $\boldsymbol{Z}$*, iff

$$P(\boldsymbol{X}, \boldsymbol{Y} \mid \boldsymbol{Z}) \ = \ P(\boldsymbol{X} \mid \boldsymbol{Z}) \cdot P(\boldsymbol{Y} \mid \boldsymbol{Z}) \Leftrightarrow P(\boldsymbol{X} \mid \boldsymbol{Y}, \boldsymbol{Z}) \ = \ P(\boldsymbol{X} \mid \boldsymbol{Z})$$

Shorthand notation: $\boldsymbol{X} \perp\!\!\!\perp_P \boldsymbol{Y} \mid \boldsymbol{Z}$

Conditional Independence

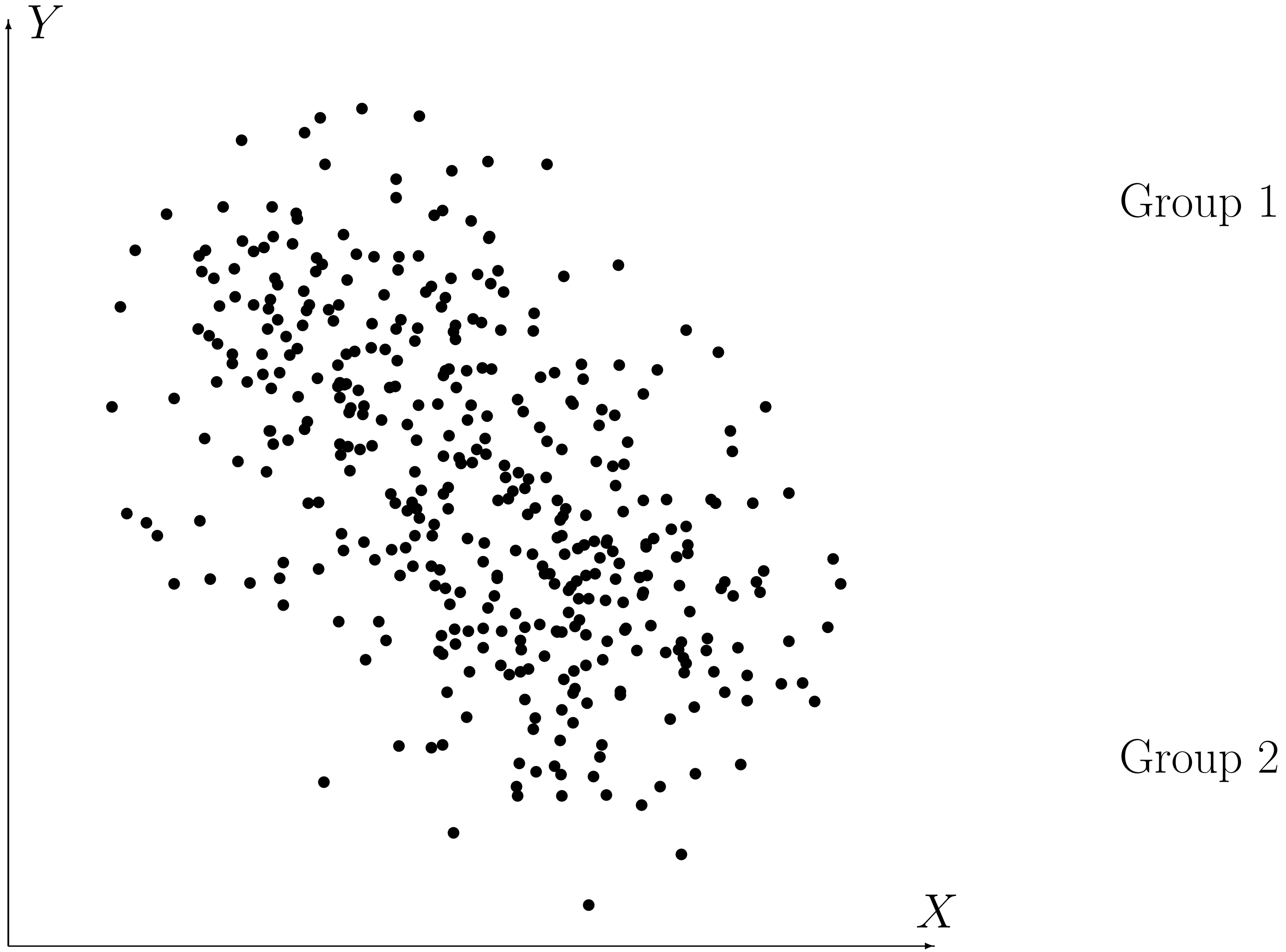
The complete condition for $\boldsymbol{X} \perp\!\!\!\perp_P \boldsymbol{Y} \mid \boldsymbol{Z}$ would read as follows:

$$
\begin{aligned}
&\forall a_1 \in \mathrm{dom}(A_1) : \cdots \forall a_k \in \mathrm{dom}(A_k) : \\
&\quad \forall b_1 \in \mathrm{dom}(B_1) : \cdots \forall b_l \in \mathrm{dom}(B_l) : \\
&\qquad \forall c_1 \in \mathrm{dom}(C_1) : \cdots \forall c_m \in \mathrm{dom}(C_m) : \\
&\qquad\quad P(A_1 = a_1, \ldots, A_k = a_k, B_1 = b_1, \ldots, B_l = b_l \mid C_1 = c_1, \ldots, C_m = c_m) \\
&\qquad\quad\quad = P(A_1 = a_1, \ldots, A_k = a_k \mid C_1 = c_1, \ldots, C_m = c_m) \\
&\qquad\quad\quad\ \cdot P(B_1 = b_1, \ldots, B_l = b_l \mid C_1 = c_1, \ldots, C_m = c_m)
\end{aligned}
$$

Remarks:

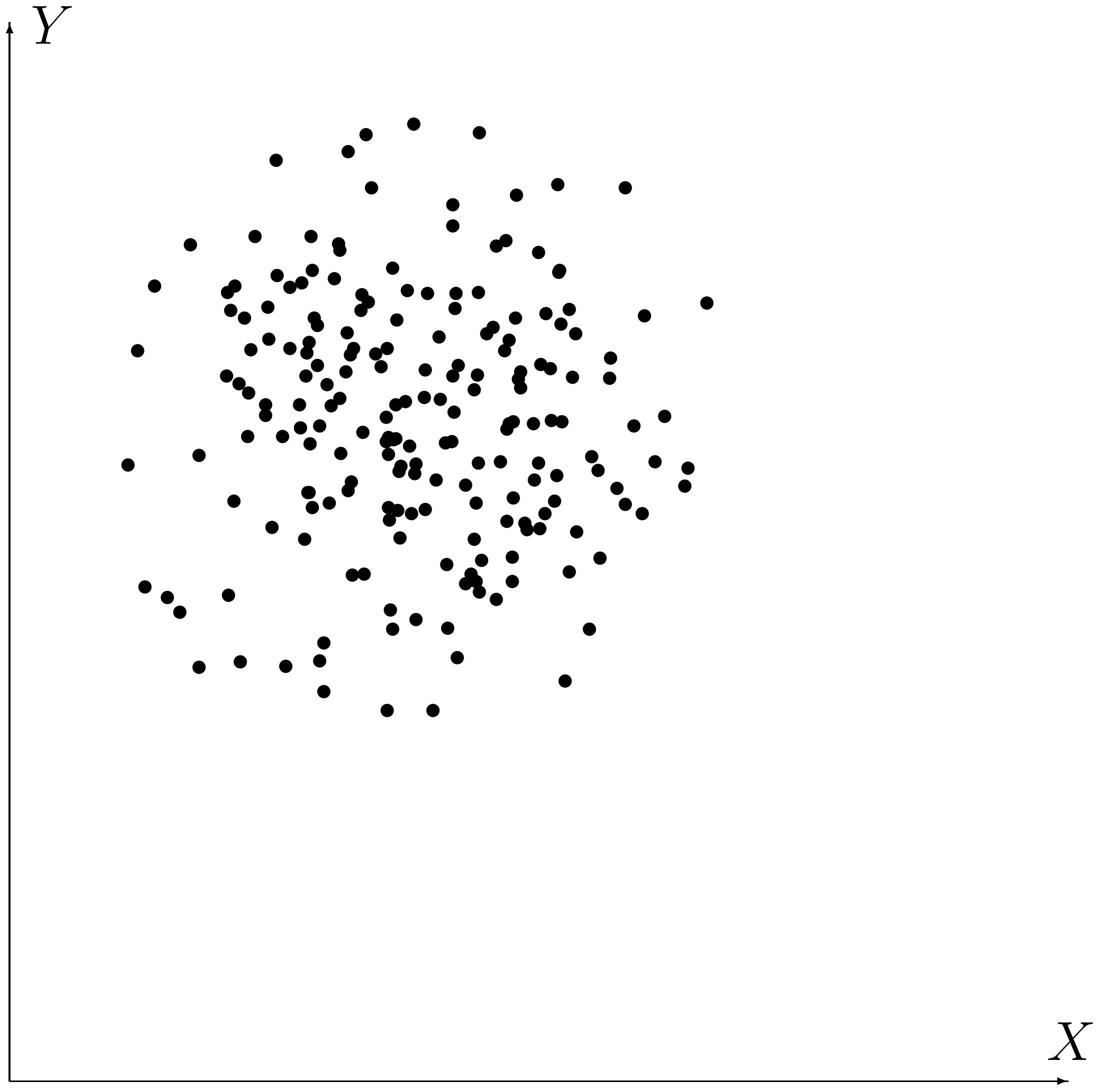
1. If $\boldsymbol{Z} = \emptyset$ we get (unconditional) independence.
2. We do not use curly braces ($\{\}$) for the sets if the context is clear. Likewise, we use X instead of $\boldsymbol{X}$ to denote sets.

Conditional Independence — Example 1



(Weak) Dependence in the entire dataset: X and Y dependent.

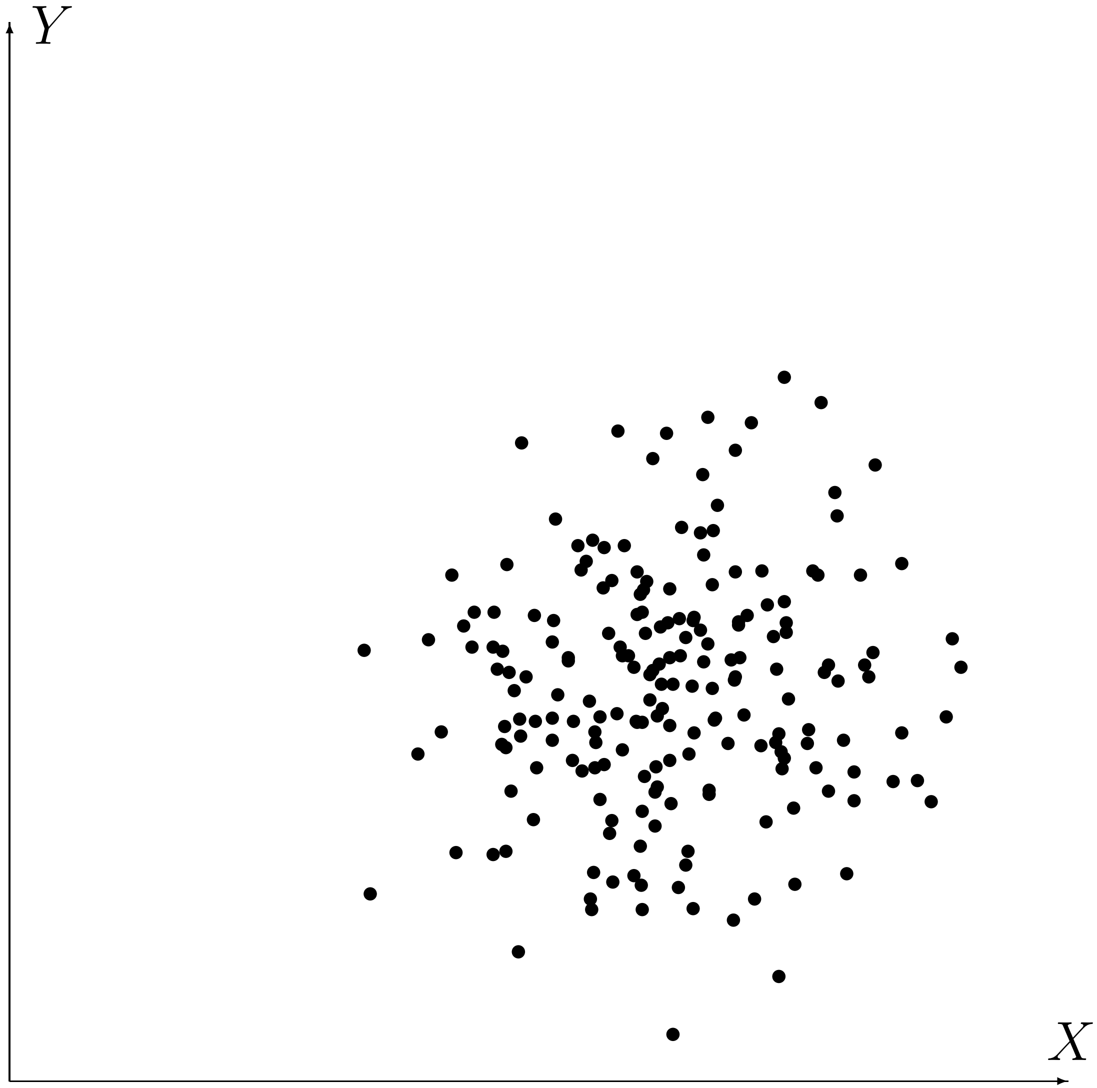
Conditional Independence — Example 1



Group 1

No Dependence in Group 1: X and Y conditionally independent given Group 1.

Conditional Independence — Example 1



No Dependence in Group 2: X and Y conditionally independent given Group 2.

Conditional Independence — Example 2

- $\mathrm{dom}(G) = \{\mathsf{mal}, \mathsf{fem}\}$ Geschlecht (gender)
- $\mathrm{dom}(S) = \{\mathsf{sm}, \overline{\mathsf{sm}}\}$ Raucher (smoker)
- $\mathrm{dom}(M) = \{\mathsf{mar}, \overline{\mathsf{mar}}\}$ Verheiratet (married)
- $\mathrm{dom}(P) = \{\mathsf{preg}, \overline{\mathsf{preg}}\}$ Schwanger (pregnant)

p_{GSMP}		G = mal		G = fem	
		S = sm	S = $\overline{\text{sm}}$	S = sm	S = $\overline{\text{sm}}$
M = mar	P = preg	0	0	0.01	0.05
	P = $\overline{\text{preg}}$	0.04	0.16	0.02	0.12
M = $\overline{\text{mar}}$	P = preg	0	0	0.01	0.01
	P = $\overline{\text{preg}}$	0.10	0.20	0.07	0.21

Conditional Independence — Example 2

$$\begin{aligned} P(\mathsf{G}{=}\mathsf{fem}) &= P(\mathsf{G}{=}\mathsf{mal}) = 0.5 & \qquad P(\mathsf{P}{=}\mathsf{preg}) &= 0.08 \\ P(\mathsf{S}{=}\mathsf{sm}) &= 0.25 & \qquad P(\mathsf{M}{=}\mathsf{mar}) &= 0.4 \end{aligned}$$

Gender and Smoker are not independent:

$$P(\mathsf{G}{=}\mathsf{fem} \mid \mathsf{S}{=}\mathsf{sm}) = 0.44 \neq 0.5 = P(\mathsf{G}{=}\mathsf{fem})$$

Gender and Marriage are marginally independent but conditionally dependent given Pregnancy:

$$P(\mathsf{fem}, \mathsf{mar} \mid \overline{\mathsf{preg}}) \approx 0.152 \quad \neq \quad 0.169 \approx P(\mathsf{fem} \mid \overline{\mathsf{preg}}) \cdot P(\mathsf{mar} \mid \overline{\mathsf{preg}})$$

Bayes Theorem

Product Rule (for events A and B):

$$P(A \cap B) = P(A \mid B)P(B) \quad \text{and} \quad P(A \cap B) = P(B \mid A)P(A)$$

Equating the right-hand sides:

$$P(A \mid B) \;=\; \frac{P(B \mid A)P(A)}{P(B)}$$

For random variables X and Y:

$$\forall x \forall y: \quad P(Y = y \mid X = x) \;=\; \frac{P(X = x \mid Y = y)P(Y = y)}{P(X = x)}$$

Generalization concerning background knowledge/evidence E:

$$P(Y \mid X, E) \;=\; \frac{P(X \mid Y, E)P(Y \mid E)}{P(X \mid E)}$$

Bayes Theorem — Application

$$P(\mathsf{toothache} \mid \mathsf{cavity}) = 0.4$$
$$P(\mathsf{cavity}) = 0.1 \qquad P(\mathsf{cavity} \mid \mathsf{toothache}) = \frac{0.4 \cdot 0.1}{0.05} = 0.8$$
$$P(\mathsf{toothache}) = 0.05$$

Why not estimate $P(\mathsf{cavity} \mid \mathsf{toothache})$ right from the start?

Causal knowledge like $P(\mathsf{toothache} \mid \mathsf{cavity})$ is more robust than diagnostic knowledge $P(\mathsf{cavity} \mid \mathsf{toothache})$.

The causality $P(\mathsf{toothache} \mid \mathsf{cavity})$ is independent of the a priori probabilities $P(\mathsf{toothache})$ and $P(\mathsf{cavity})$.

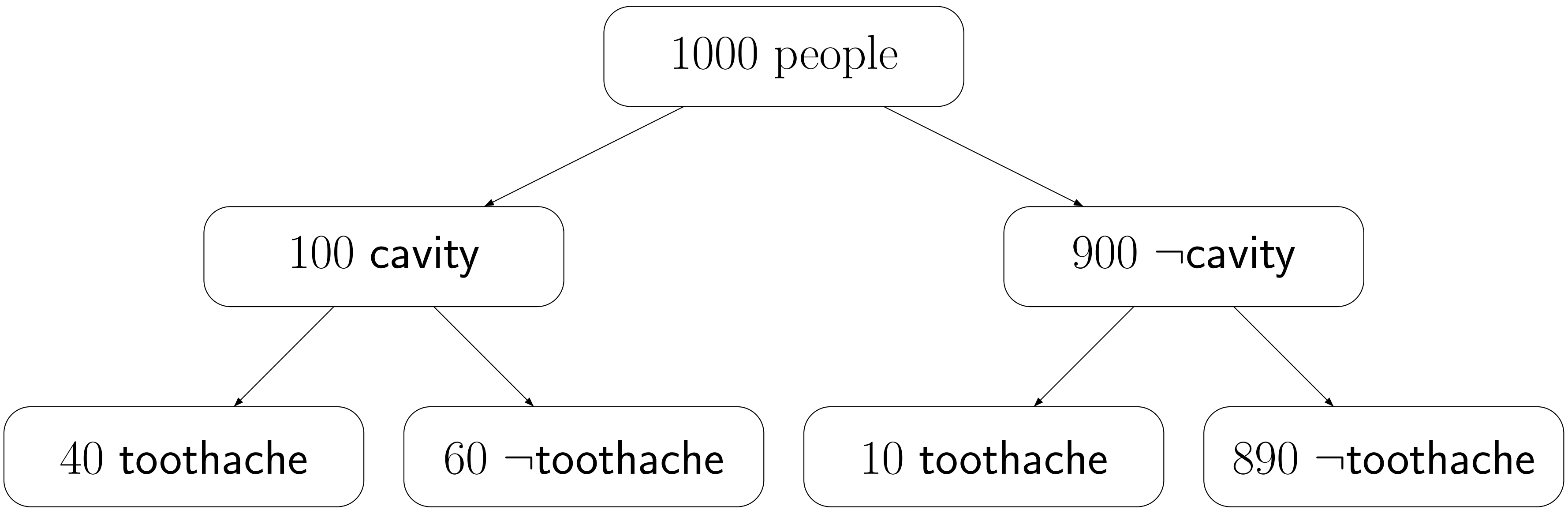
If $P(\mathsf{cavity})$ rose in a caries epidemic, the causality $P(\mathsf{toothache} \mid \mathsf{cavity})$ would remain constant whereas both $P(\mathsf{cavity} \mid \mathsf{toothache})$ and $P(\mathsf{toothache})$ would increase according to $P(\mathsf{cavity})$.

A physician, after having estimated $P(\mathsf{cavity} \mid \mathsf{toothache})$, would not know a rule for updating.

Bayes Theorem — Using absolute Numbers

$P(\text{toothache} \mid \text{cavity}) = 0.4$ $\qquad$ $P(\text{cavity}) = 0.1$

$P(\text{toothache} \mid \neg\text{cavity}) = \frac{1}{90}$ $\qquad$ $P(\text{cavity} \mid \text{toothache}) = \dfrac{40}{40 + 10} = 0.8$



$$P(C \mid T) = \frac{P(T \mid C) \cdot P(C)}{P(T)} = \frac{P(T \mid C) \cdot P(C)}{P(T \mid C) \cdot P(C) + P(T \mid \neg C) \cdot P(\neg C)}$$

Relative Probabilities

Assumption:
We would like to consider the probability of the diagnosis $\mathsf{GumDisease}$ as well.

$$\begin{aligned} P(\mathsf{toothache} \mid \mathsf{gumdisease}) &= 0.7 \\ P(\mathsf{gumdisease}) &= 0.02 \end{aligned}$$

Which diagnosis is more probable?

If we are interested in *relative probabilities* only (which may be sufficient for some decisions), $P(\mathsf{toothache})$ needs not to be estimated:

$$\begin{aligned} \frac{P(C \mid T)}{P(G \mid T)} &= \frac{P(T \mid C)P(C)}{P(T)} \cdot \frac{P(T)}{P(T \mid G)P(G)} \\ &= \frac{P(T \mid C)P(C)}{P(T \mid G)P(G)} = \frac{0.4 \cdot 0.1}{0.7 \cdot 0.02} \\ &= 28.57 \end{aligned}$$

Normalization

If we are interested in the absolute probability of $P(C \mid T)$ but do not know $P(T)$, we may conduct a complete case analysis (according C) and exploit the fact that $P(C \mid T) + P(\neg C \mid T) = 1$.

$$\begin{aligned}
P(C \mid T) &= \frac{P(T \mid C)P(C)}{P(T)} \\
P(\neg C \mid T) &= \frac{P(T \mid \neg C)P(\neg C)}{P(T)} \\
1 = P(C \mid T) + P(\neg C \mid T) &= \frac{P(T \mid C)P(C)}{P(T)} + \frac{P(T \mid \neg C)P(\neg C)}{P(T)} \\
P(T) &= P(T \mid C)P(C) + P(T \mid \neg C)P(\neg C)
\end{aligned}$$

Normalization

Plugging into the equation for $P(C \mid T)$ yields:

$$P(C \mid T) = \frac{P(T \mid C)P(C)}{P(T \mid C)P(C) + P(T \mid \neg C)P(\neg C)}$$

For general random variables, the equation reads:

$$P(Y = y \mid X = x) = \frac{P(X = x \mid Y = y)P(Y = y)}{\sum_{\forall y' \in \mathrm{dom}(Y)} P(X = x \mid Y = y')P(Y = y')}$$

Note the "loop variable" y'. Do not confuse with y.

Multiple Evidences

The patient complains about a toothache. From this first evidence the dentist infers:

$$P(\mathsf{cavity} \mid \mathsf{toothache}) = 0.8$$

The dentist palpates the tooth with a metal probe which catches into a fracture:

$$P(\mathsf{cavity} \mid \mathsf{fracture}) = 0.95$$

Both conclusions might be inferred via Bayes rule. But what does the combined evidence yield? Using Bayes rule further, the dentist might want to determine:

$$P(\mathsf{cavity} \mid \mathsf{toothache} \wedge \mathsf{fracture}) = \frac{P(\mathsf{toothache} \wedge \mathsf{fracture} \mid \mathsf{cavity}) \cdot P(\mathsf{cavity})}{P(\mathsf{toothache} \wedge \mathsf{fracture})}$$

Multiple Evidences

Problem:
He needs $P(\textsf{toothache} \wedge \textsf{catch} \mid \textsf{cavity})$, i. e. diagnostics knowledge for all combinations of symptoms in general. Better incorporate evidences step-by-step:

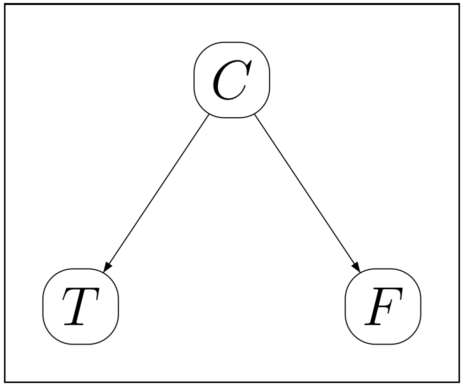
$$P(Y \mid X, E) = \frac{P(X \mid Y, E) P(Y \mid E)}{P(X \mid E)}$$

Abbreviations:

C — cavity

T — toothache

F — fracture



Objective:
Computing $P(C \mid T, F)$ with just using information about $P(\cdot \mid C)$ and under exploitation of independence relations among the variables.

Multiple Evidences

A priori: $P(C)$

Evidence toothache: $P(C \mid T) \quad = P(C)\, \dfrac{P(T \mid C)}{P(T)}$

Evidence fracture: $P(C \mid T, F) \quad = P(C \mid T)\, \dfrac{P(F \mid C, T)}{P(F \mid T)}$

Information about conditional independence

$$P(F \mid C, T) = P(F \mid C)$$

$$P(C \mid T, F) \;=\; P(C)\, \frac{P(T \mid C)}{P(T)}\, \frac{P(F \mid C)}{P(F \mid T)}$$

Seems that we still have to cope with symptom inter-dependencies?!

Multiple Evidences

Compound equation from last slide:

$$P(C \mid T, F) \;=\; P(C)\,\frac{P(T \mid C)\,P(F \mid C)}{P(T)\,P(F \mid T)}$$

$$\;=\; P(C)\,\frac{P(T \mid C)\,P(F \mid C)}{P(F, T)}$$

$P(F, T)$ is a normalizing constant and can be computed if $P(F \mid \neg C)$ and $P(T \mid \neg C)$ are known:

$$P(F, T) = \underbrace{P(F, T \mid C)}_{P(F|C)P(T|C)}\,P(C) + \underbrace{P(F, T \mid \neg C)}_{P(F|\neg C)P(T|\neg C)}\,P(\neg C)$$

Therefore, we finally arrive at the following solution...

Multiple Evidences

$$P(C \mid F,T) = \frac{\boxed{P(C)}\ \boxed{P(T \mid C)}\ \boxed{P(F \mid C)}}{\boxed{P(F \mid C)}\ \boxed{P(T \mid C)}\ \boxed{P(C)} + \boxed{P(F \mid \neg C)}\ \boxed{P(T \mid \neg C)}\ \boxed{P(\neg C)}}$$

Note that we only use causal probabilities $P(\ \cdot \mid C)$ together with the a priori (marginal) probabilities $P(C)$ and $P(\neg C)$.

Multiple Evidences — Summary

Multiple evidences can be treated by reduction on

- a priori probabilities
- (causal) conditional probabilities for the evidence
- under assumption of conditional independence

General rule:

$$P(Z \mid X, Y) \; = \; \alpha \; P(Z) \; P(X \mid Z) \; P(Y \mid Z)$$

for X and Y conditionally independent given Z and with normalizing constant α.

Monty Hall Puzzle

Marylin Vos Savant in her riddle column in the New York Times:

You are a candidate in a game show and have to choose between three doors. Behind one of them is a Porsche, whereas behind the other two there are goats. After you chose a door, the host Monty Hall (who knows what is behind each door) opens another (not your chosen one) door with a goat. Now you have the choice between keeping your chosen door or choose the remaining one.

Which decision yields the best chance of winning the Porsche?

Monty Hall Puzzle

G	You win the Porsche.
R	You revise your decision.
A	Behind your initially chosen door is (and remains) the Porsche.

$$\begin{aligned}
P(G \mid R) &= P(G, A \mid R) + P(G, \overline{A} \mid R) \\
&= P(G \mid A, R)P(A \mid R) + P(G \mid \overline{A}, R)P(\overline{A} \mid R) \\
&= 0 \cdot P(A \mid R) + 1 \cdot P(\overline{A} \mid R) \\
&= P(\overline{A} \mid R) = P(\overline{A}) = \frac{2}{3}
\end{aligned}$$

$$\begin{aligned}
P(G \mid \overline{R}) &= P(G, A \mid \overline{R}) + P(G, \overline{A} \mid \overline{R}) \\
&= P(G \mid A, \overline{R})P(A \mid \overline{R}) + P(G \mid \overline{A}, \overline{R})P(\overline{A} \mid \overline{R}) \\
&= 1 \cdot P(A \mid \overline{R}) + 0 \cdot P(\overline{A} \mid \overline{R}) \\
&= P(A \mid \overline{R}) = P(A) = \frac{1}{3}
\end{aligned}$$

Simpson's Paradox

Example: C = Patient takes medication, E = patient recovers

	E	$\neg E$	$\sum$	Recovery rate
C	20	20	40	50%
$\neg C$	16	24	40	40%
$\sum$	36	44	80	

Men	E	$\neg E$	$\sum$	Rec.rate	Women	E	$\neg E$	$\sum$	Rec.rate
C	18	12	30	60%	C	2	8	10	20%
$\neg C$	7	3	10	70%	$\neg C$	9	21	30	30%
	25	15	40			11	29	40	

$$
\begin{array}{rrcl}
 & P(E \mid C) & > & P(E \mid \neg C) \\
\text{but} & P(E \mid C, M) & < & P(E \mid \neg C, M) \\
 & P(E \mid C, W) & < & P(E \mid \neg C, W)
\end{array}
$$

Excursus: Focusing vs. Revision

Philosophical topic, studied e.g. by Kant, Gärdenfors

Example for Focusing

- Prior knowledge: a fair dice
- New evidence: the result is an odd number
- Aposteriori knowledge via focusing: conditional probability
- Underlying probability measure did not change

Example for Revision

- Prior knowledge: a fair dice
- New evidence: weight near the 6
- Belief change via revision
- Underlying probability measure did change

Excursus: Causality vs. Correlation

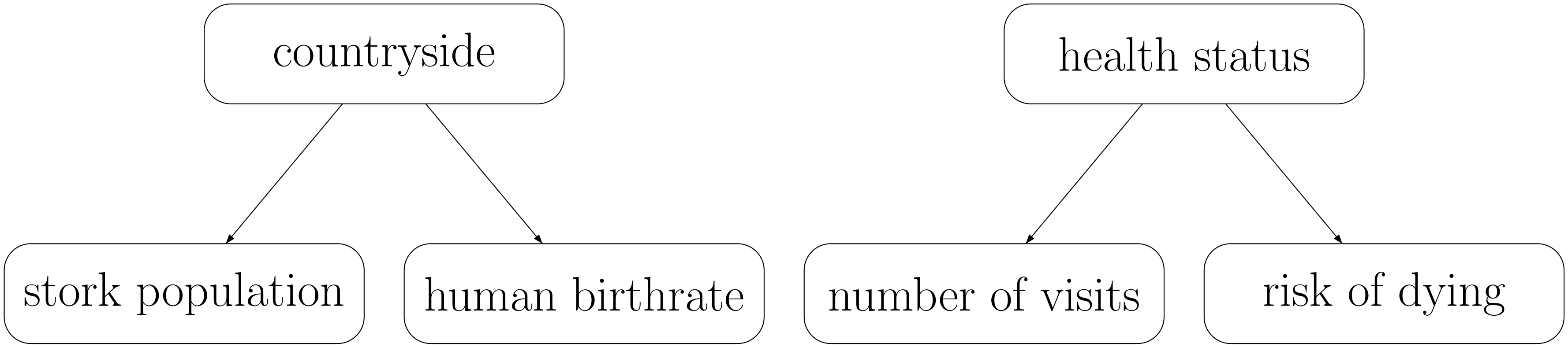
Philosophical topic, studied e.g. by Aristoteles, still under discussion

Press acceleration pedal → car is faster (causality)

Stork population high → human birthrate (correlation, but no causality)

Visit doctor often → high risk of dying (correlation, but no causality)

Explanation of the correlations by using additional attributes:



Probabilistic Reasoning

Probabilistic reasoning is difficult and may be problematic:

- $P(A \wedge B)$ is not determined simply by $P(A)$ and $P(B)$:
  $P(A) = P(B) = 0.5 \quad \Rightarrow \quad P(A \wedge B) \in [0, 0.5]$
- $P(C \mid A) = x, P(C \mid B) = y \quad \Rightarrow \quad P(C \mid A \wedge B) \in [0, 1]$
  Probabilistic logic is *not truth functional*!

Central problem: How does additional information affect the current knowledge?
I. e., if $P(B \mid A)$ is known, what can be said about $P(B \mid A \wedge C)$?

High complexity: n propositions $\rightarrow 2^n$ full conjunctives

Hard to specify these probabilities.

Summary

Uncertainty is inevitable in complex and dynamic scenarios that force agents to cope with ignorance.

Probabilities express the agent's inability to vote for a definitive decision. They model the degree of belief.

If an agent violates the axioms of probability, it may exhibit irrational behavior in certain circumstances.

The Bayes rule is used to derive unknown probabilities from present knowledge and new evidence.

Multiple evidences can be effectively included into computations exploiting conditional independencies.

Decomposition